



This item has a late penalty. A penalty of 10% will be applied to your score per late day after the deadline. [Learn More](#)

Programming Assignment: Implementing the Decision Tree Algorithm

✓ Passed · 98/100 points

Deadline The assignment was due on May 5, 09:59 PM PDT
You can still pass this assignment before the course ends.

Instructions

My submission

Discussions

In this assignment, you will build a classification framework. You need to implement the decision tree algorithm with Gini-index as the attribute selection measure.

You will be given a training dataset "training.txt" and a testing dataset "testing.txt". In the training dataset, the data format is

<label> <index1>:<value1> <index2>:<value2> ...

.....

Each line contains an instance and is ended by a '\n' character. <label> is an integer indicating the class label. The pair <index>:<value> gives a feature (attribute) value: <index> is a non-negative integer and <value> is a number (we only consider categorical attributes in this assignment). Note that one attribute may have more than 2 possible values, meaning it is a multi-value categorical attribute.

In the testing dataset, the data format is

<index1>:<value1> <index2>:<value2> ...



.....



You will no longer have the <label> in each line.

training.txt

testing.txt

You need to submit a file titled "result.txt". Each line contains one integer representing the predicted label of a testing sample.

You will be graded based on whether your file format is correct and on the precision of your classifier on the testing dataset. You will get a full score as long as your precision is above a certain threshold.

How to submit

When you're ready to submit, you can upload files for each part of the assignment on the "My submission" tab.

